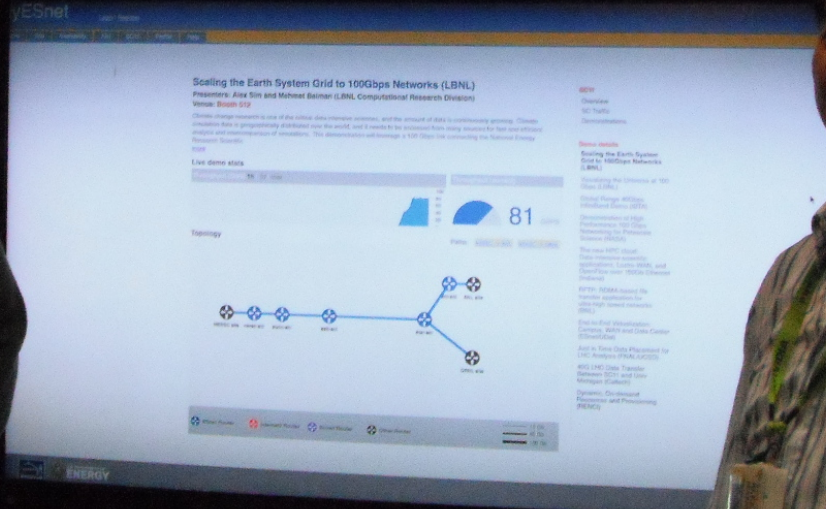
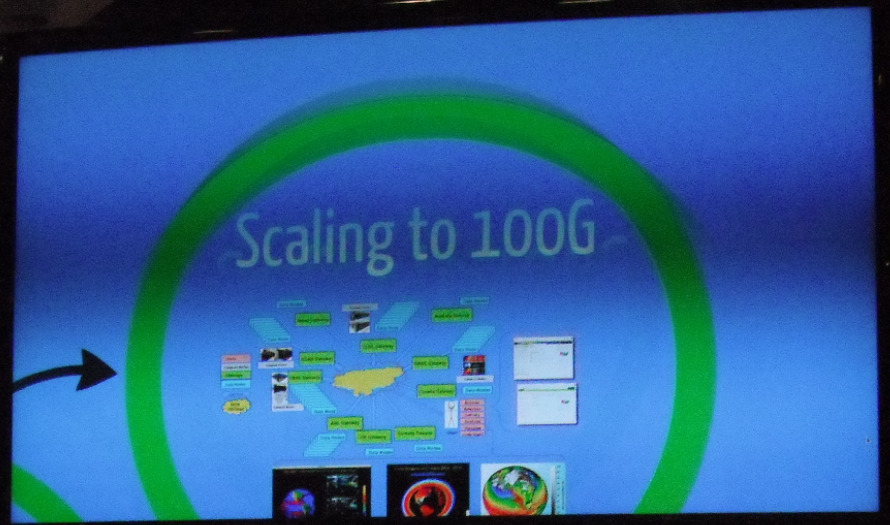


Data Movement over 100Gbps

Scaling the Earth System Grid to 100Gbps Network

Preliminary Results from SC'11 100Gbps demo



of Network

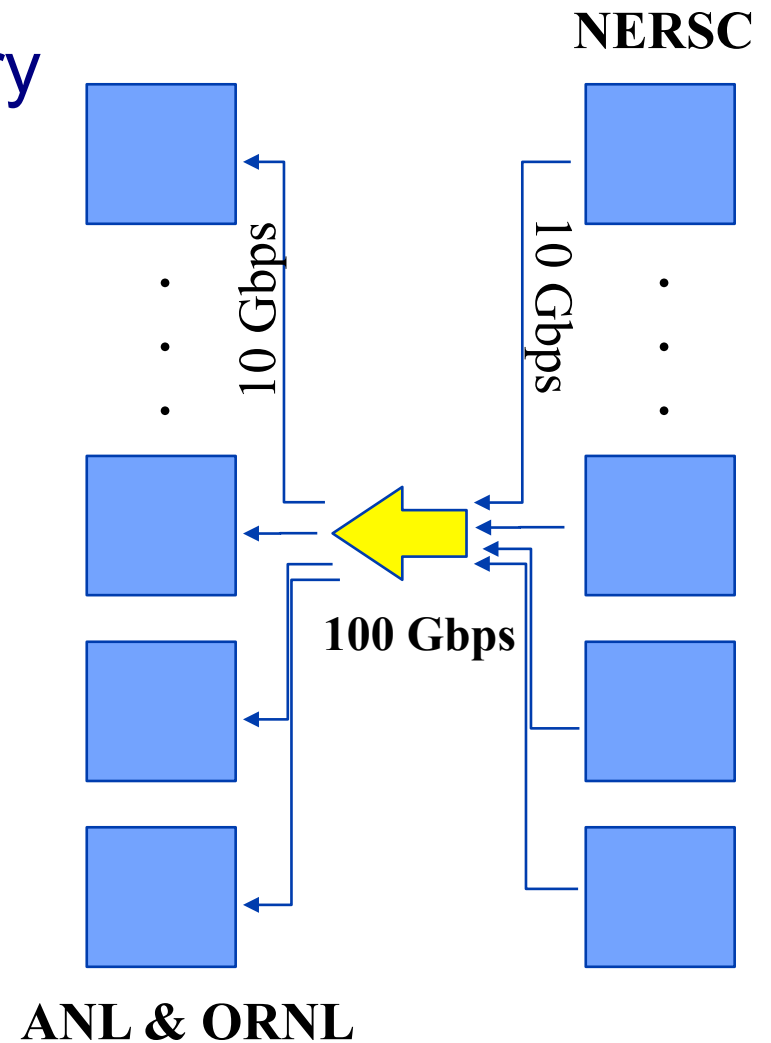
BERKELEY LAB

NERSC

Powered by

SC11 demo

- NERSC disk to ANL/ORNL memory
 - ~35TB of entire CMIP-3 dataset
 - Each node has 10Gbps connection
 - Over TCP
 - No TCP tuning (rely on system wide parameters)
 - 4MB block size
 - 2G virtual file size (cache)
 - 8 streams for each point-to-point connection



Scaling the Earth System Grid to 100Gbps Networks (LBNL)

Presenters: Alex Sim and Mehmet Balman (LBNL Computational Research Division)

Venue: Booth 512

Climate change research is one of the critical data intensive sciences, and the amount of data is continuously growing. Climate simulation data is geographically distributed over the world, and it needs to be accessed from many sources for fast and efficient analysis and intercomparison of simulations. This demonstration will leverage a 100 Gbps link connecting the National Energy Research Scientific

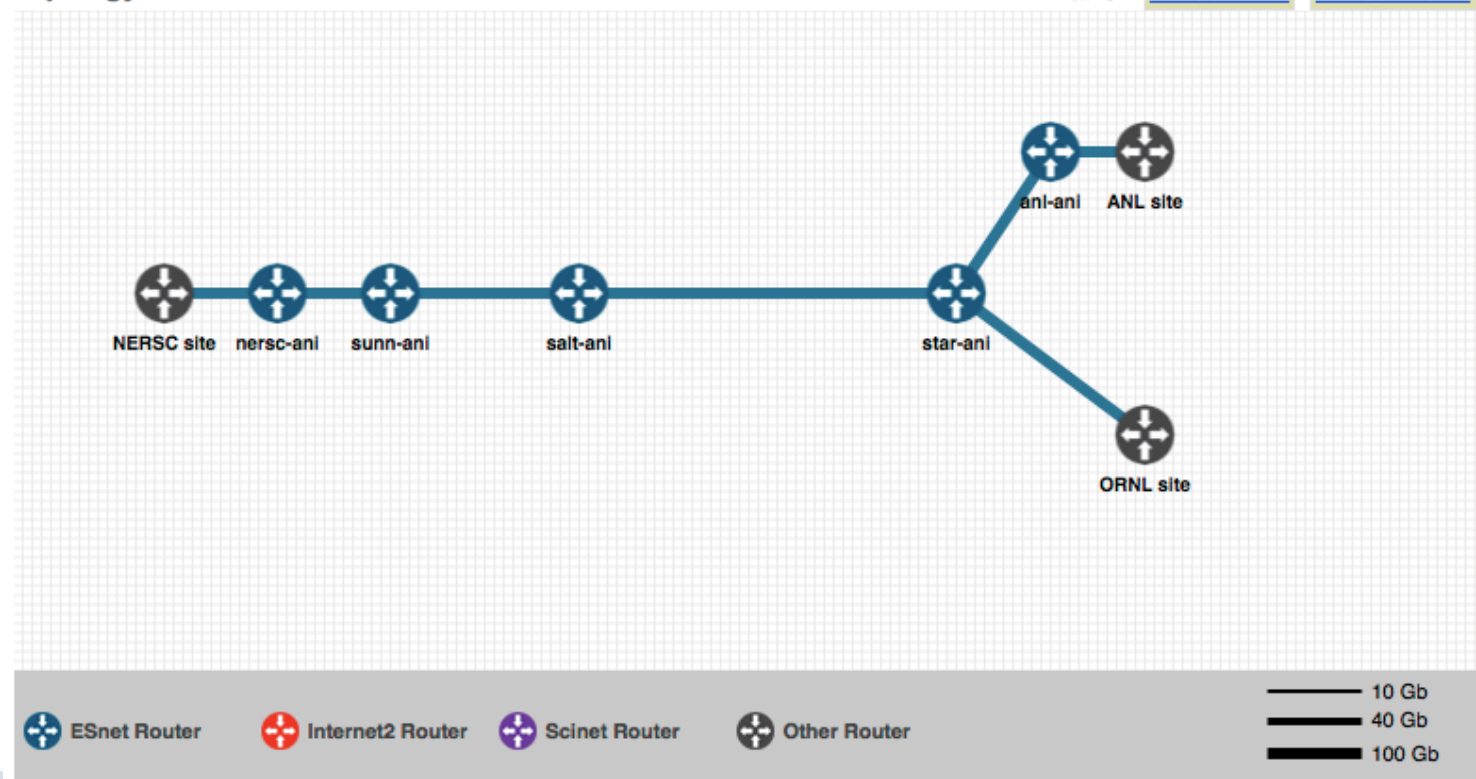
[more](#)

Live demo stats



Topology

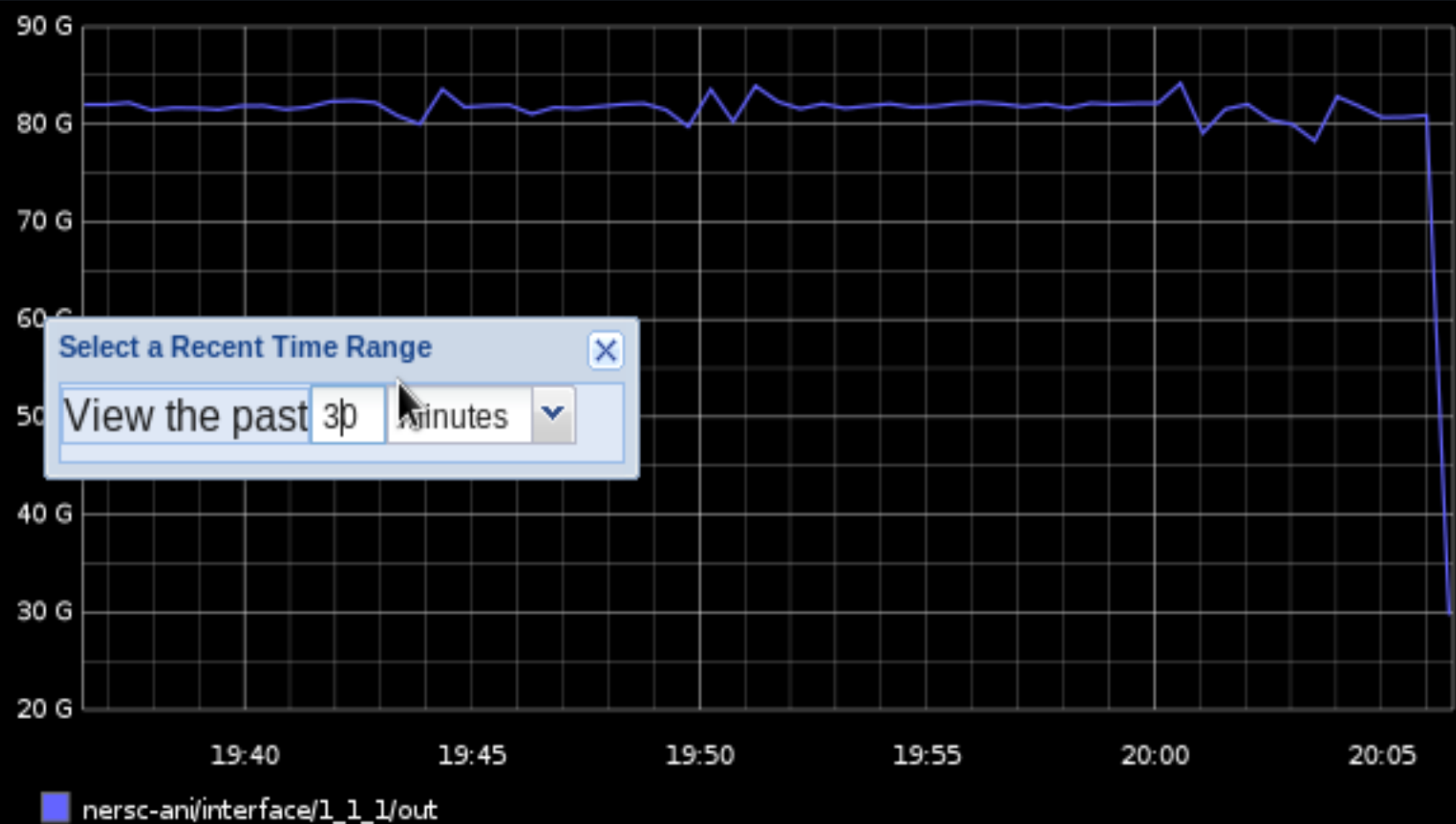
Paths: [NERSC -> ANL](#) [NERSC -> ORNL](#)



Graphite Composer



Now showing the past 30 minutes



Graph Options ▾

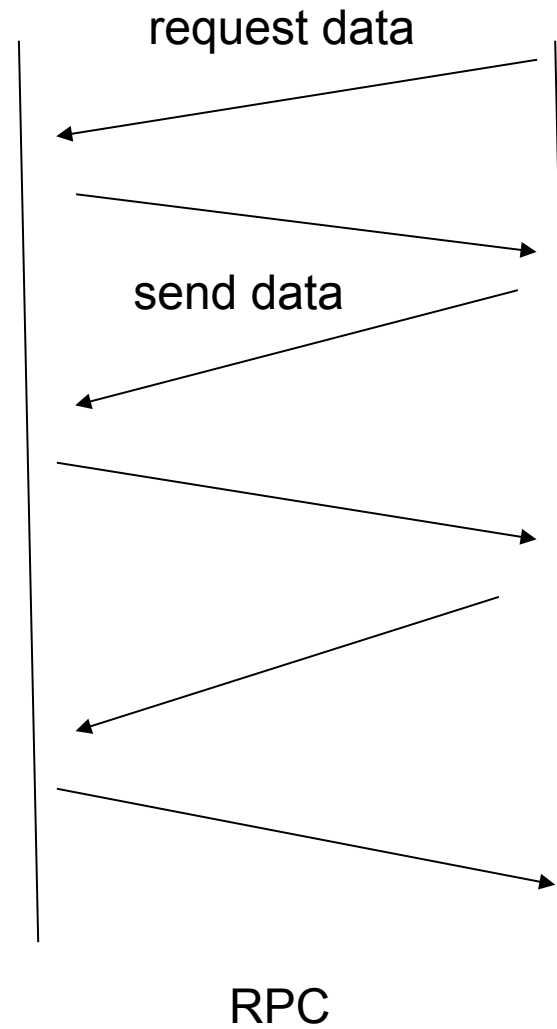
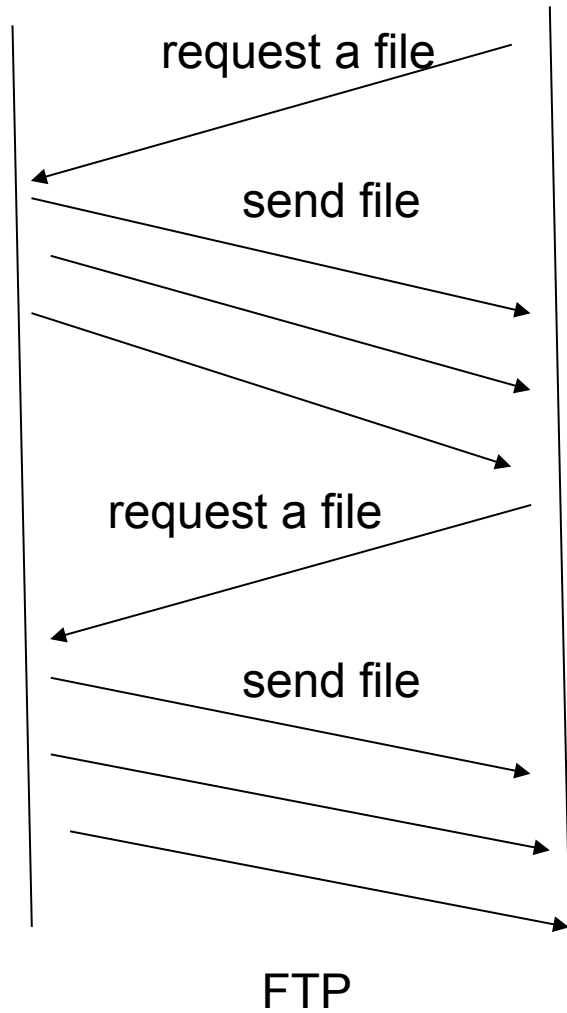
Graph Data

Auto-Refresh

Characteristics of Climate data

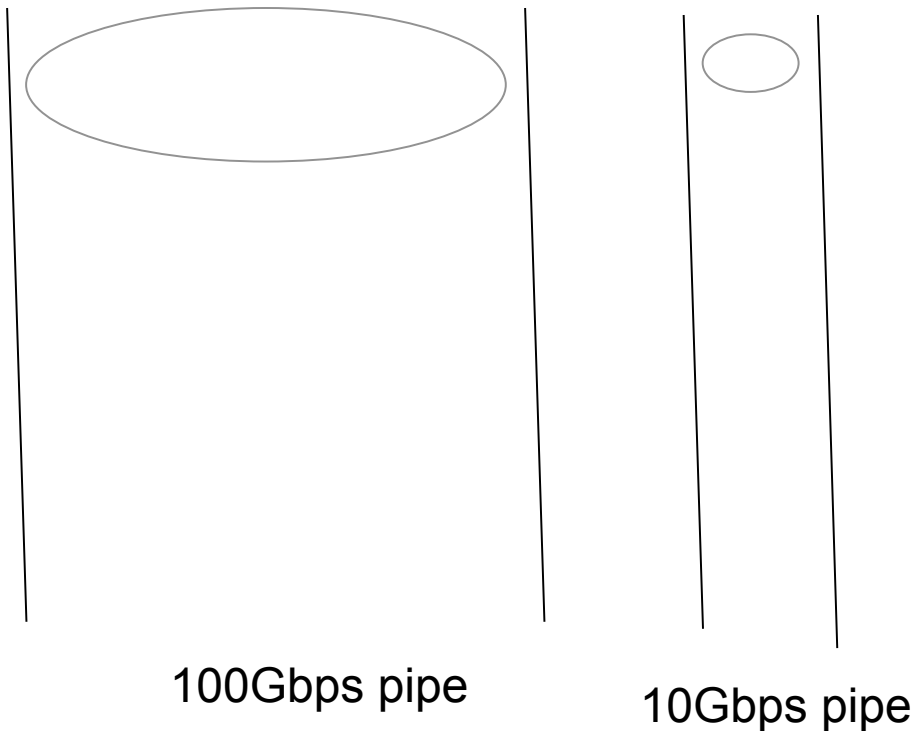
- Total size of the data is increasing
 - Requires efficient methodologies to fully utilize the underlying network infrastructure
- Many files (relatively small files)
 - In general, input data for an analysis application is structured in a set of many files
 - Are we limited to FTP?
 - Concurrent streams? (to eliminate the disadvantage of having small file sizes)

FTP ? RPC



High Bandwidth Networks

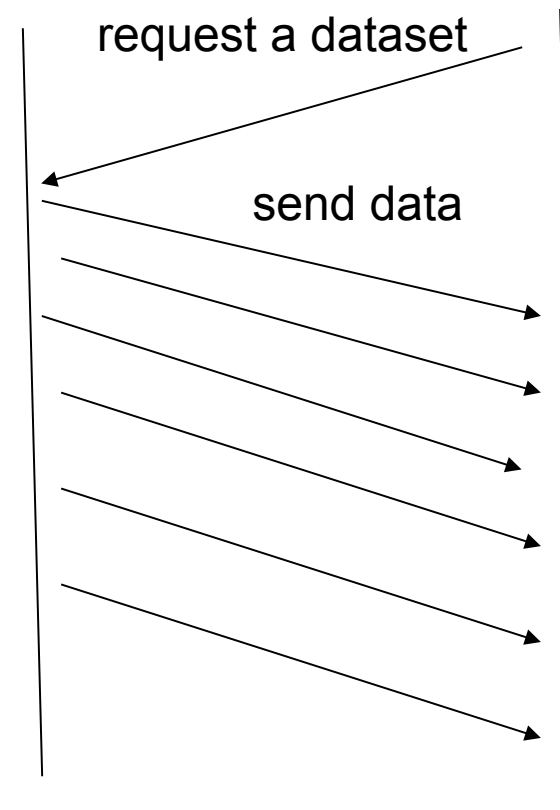
- Not necessarily high-speed (same distance)
 - Latency is still a problem



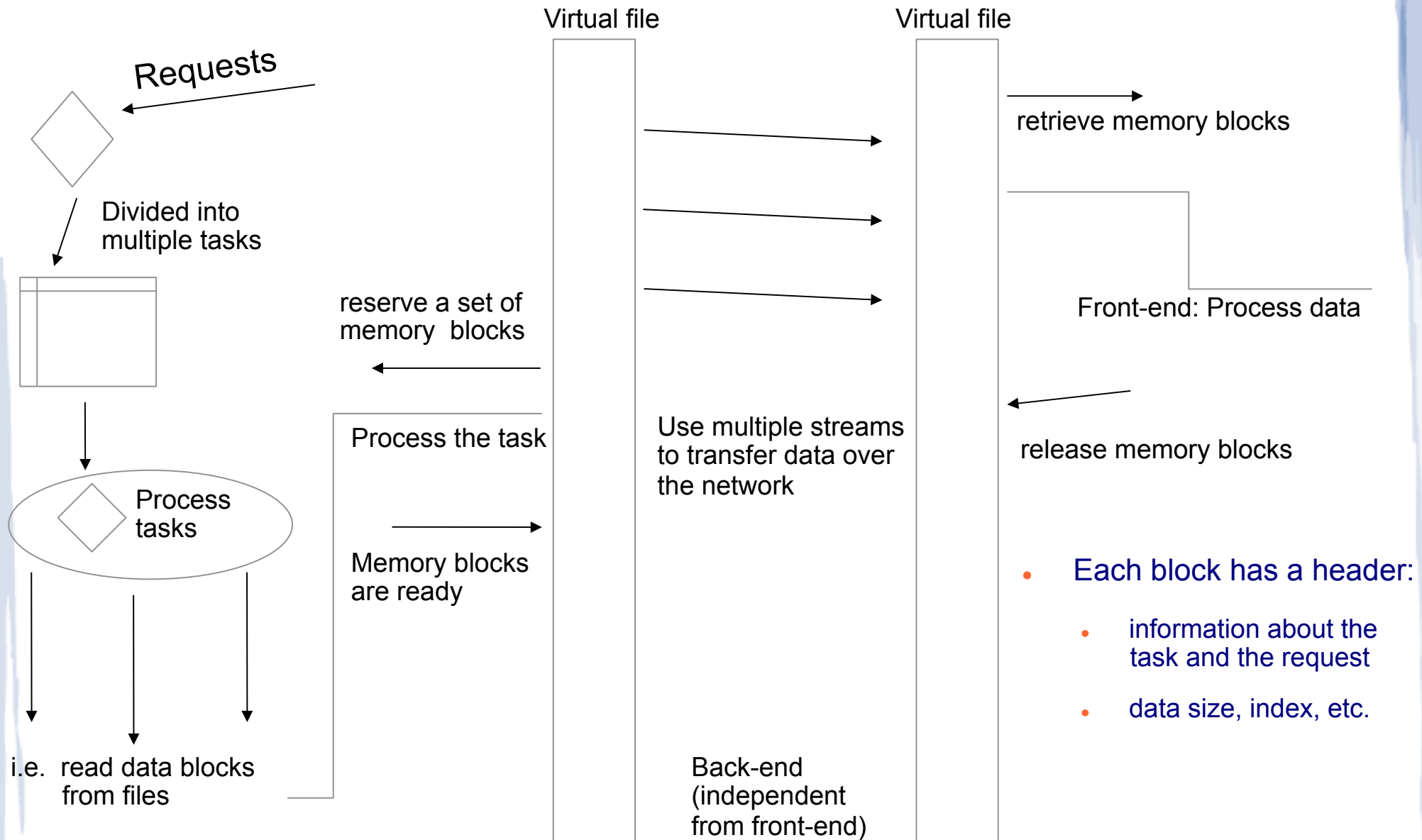
- Keep the pipe full
 - Parallelism
 - Pipelining
 - ?

Aggregate Requests

- Keep the pipe full
 - Minimize control messages
 - Aggregate requests
 - Pre-processing and post-processing at the end-nodes



ESG2NET100 tool



ESG2NET100 tool

- Virtual file is a large memory cache
 - 1G, 2G
 - Performance variations in the file system is balanced
- Back-end works independent from front-end
 - Requires more parallelism if latency is high
 - No control message is necessary in the back-end
 - Memory blocks are released once sent to remote site

End system bottleneck?

Management and tuning of multiple hosts

Number of involving host systems

Multiple streams for increased utilization

Performance monitoring in host systems

System bottleneck in end-to-end transfers

Memory overhead / CPU usage

Network is the bottleneck?

Load-balancing?

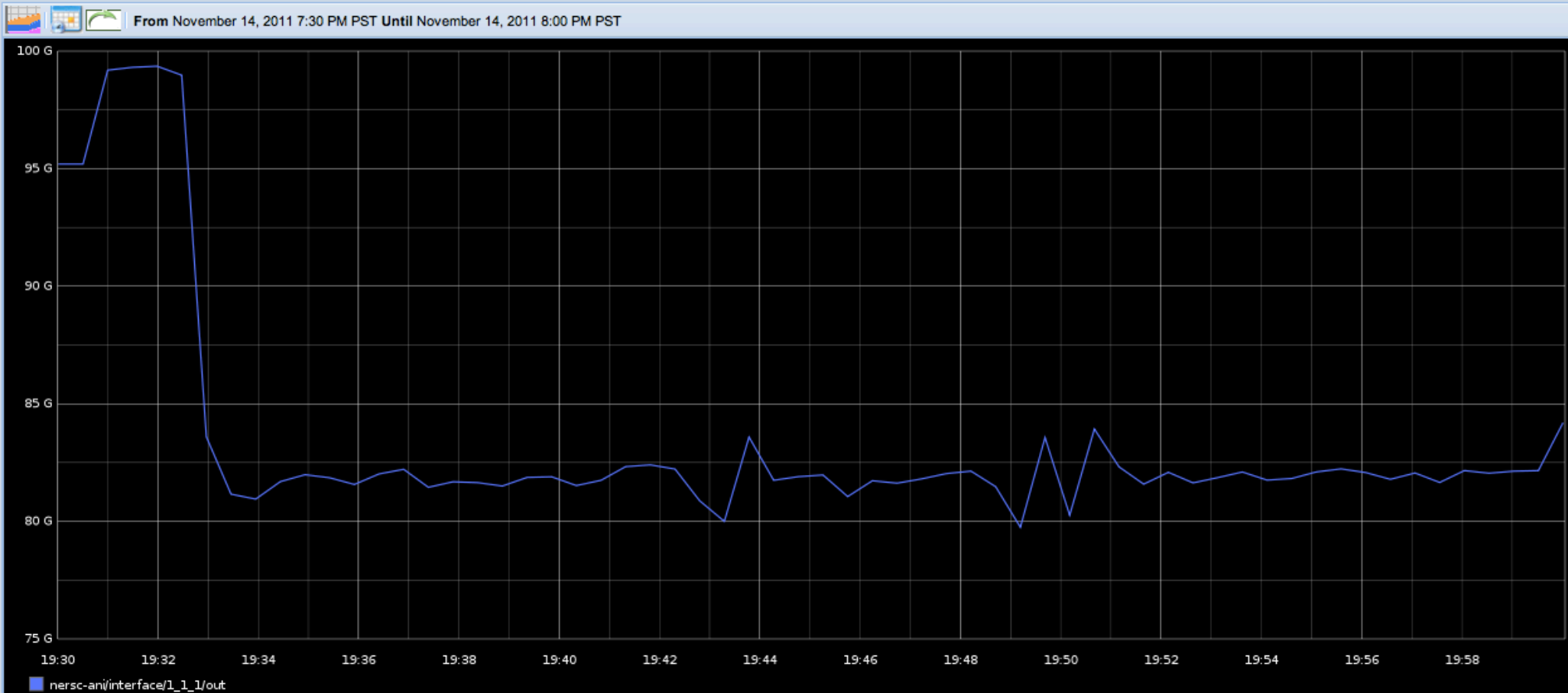


Extra

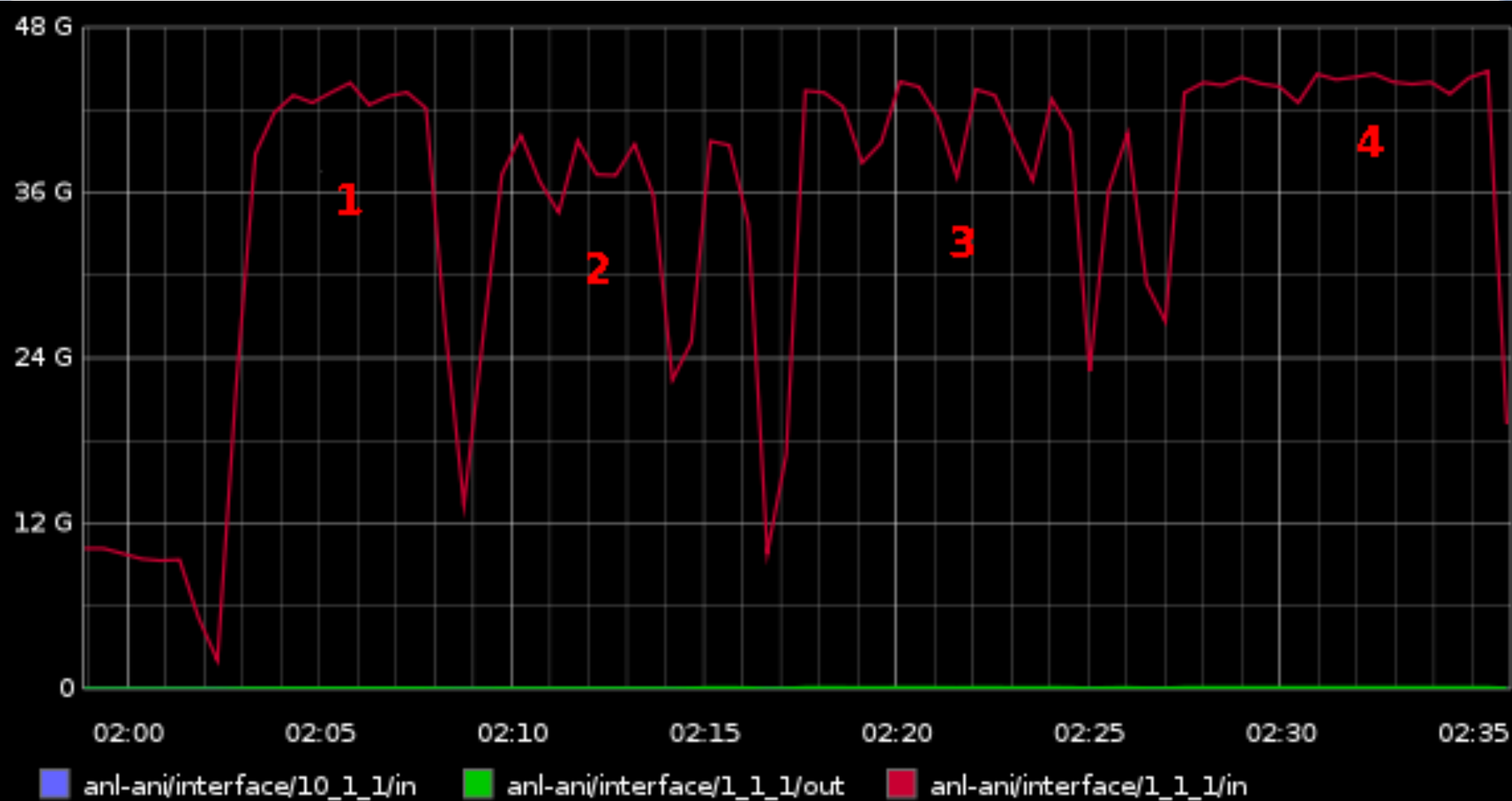
graphite

[Dashboard UI](#)
[Command-Line Interface](#)
[Login](#)
[Documentation](#)

Graphite Composer



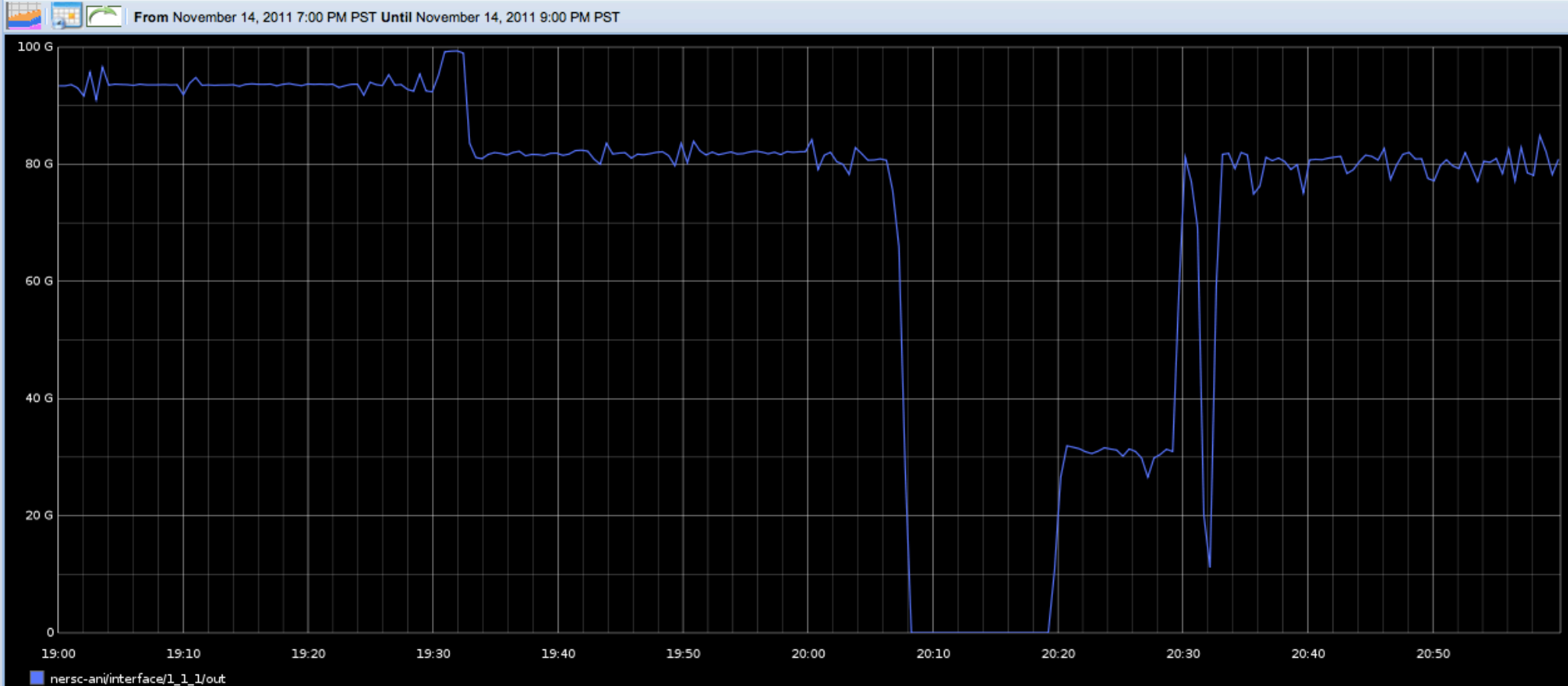
Graph Options ▾ Graph Data Auto-Refresh



graphite [Dashboard UI](#)
[Command-Line Interface](#)
[Login](#)
[Documentation](#)



Graphite Composer



Graph Options

Graph Data

Auto-Refresh